



# CREATIVITY & COPYRIGHT IN THE WORLD OF GEN-AI

*A Submission to the Government of Canada  
Consultation on Copyright in the Age of  
Generative Artificial Intelligence*

Abigail Kearney & Gareth Spanglett

January 2024

## About CIPPIC

CIPPIC is Canada's first and only public interest technology law clinic. Based at the Centre for Law, Technology and Society at the University of Ottawa's Faculty of Law, our team of legal experts and law students works together to advance the public interest on critical law and technology issues including privacy, free expression, intellectual property, telecommunications policy, and data and algorithmic governance. For more information, visit our website at [www.cippic.ca](http://www.cippic.ca).

## About the Authors

Abigail Kearney is a CIPPIC student intern and completing her final year in the English common law J.D. program at the University of Ottawa's Faculty of Law. Abigail previously completed a specialized, honours bachelor's degree in biomedical sciences and a master's degree in biomedical computing, both from Queen's University. She has experience working as a summer student in intellectual property law.

Gareth Spanglett is a CIPPIC student intern and completing his final year in the English common law J.D. program at the University of Ottawa's Faculty of Law. Gareth also works in IT security and data governance in the semiconductor industry with two decades of experience.

This document was prepared under the supervision of David Fewer, CIPPIC's Director and General Counsel.



CIPPIC has licenced this work under a [Creative Commons Attribution-ShareAlike 4.0 International Licence](https://creativecommons.org/licenses/by-sa/4.0/).

Cover image art is by Michael Dzedzic (2020) /Unsplash. Used under license: [unsplash.com/license](https://unsplash.com/license).

# Table of Contents

<b>1.</b>	<b>Introduction .....</b>	<b>1</b>
1.1.	Summary of CIPPIC’s Position.....	1
<b>2.</b>	<b>Technical Aspects of AI Development &amp; Human Involvement .....</b>	<b>2</b>
2.1.	Context .....	2
2.2.	CIPPIC’s Submissions .....	2
2.2.1.	Human Involvement in AI Development .....	2
2.2.2.	The use of Gen-AI in the Legal Field .....	2
<b>3.</b>	<b>The Impacts of Text &amp; Data Mining (TDM) on Copyrights .....</b>	<b>3</b>
3.1.	Context .....	3
3.2.	CIPPIC’s Submissions .....	3
3.2.1.	Necessary Clarification for TDM and its Impacts on Copyright in Canada .....	3
3.2.2.	The Prevalence of TDM in Canadian Industries.....	4
3.2.3.	Challenges for Copyright Holders in Licensing Use of Works for TDM Purposes .....	5
3.2.4.	Current TDM Licensing Practices and its Imposed Challenges .....	5
3.2.5.	How the Copyright Act should view Text and Data Mining .....	6
3.2.6.	Disclosure Obligations for use of Copyrighted Works in TDM and Gen-AI Training .....	6
3.2.7.	Remuneration for use of Copyrighted Works in TDM and Gen-AI Training.....	7
3.2.8.	International TDM Approaches .....	7
<b>4.</b>	<b>Authorship and Ownership of AI Generated Works .....</b>	<b>8</b>
4.1.	Context .....	8
4.2.	CIPPIC’s Submissions .....	9
4.2.1.	Uncertainties surrounding Authorship and Ownership for Gen-AI Works .....	9
4.2.2.	Amending Copyright Authorship and Ownership in light of Gen-AI.....	9
4.2.3.	International Authorship and Ownership Approaches.....	9
<b>5.</b>	<b>Liability for Infringement by Gen-AI .....</b>	<b>10</b>
5.1.	Context .....	10
5.2.	CIPPIC’s Submissions .....	11
5.2.1.	Applying the Current Infringement Framework to Infringing, AI-Generated Works .....	11
5.2.2.	Barriers to Establishing Infringement by Gen-AI: Access to the Original Work.....	11
5.2.3.	Commercial Use & Risk Mitigation for Infringing AI-Generated Works.....	12
5.2.4.	Uncertainties on Liability for Copyright Infringement by an AI-Generated Work .....	13
<b>6.</b>	<b>Additional Considerations.....</b>	<b>14</b>
6.1.	CIPPIC’s Submissions .....	14
6.1.1.	Vicarious Liability.....	14
6.1.2.	Rights Management Information .....	15
<b>7.</b>	<b>Conclusion .....</b>	<b>15</b>

# 1. Introduction

The Government of Canada has requested consultation on how it can amend the *Copyright Act* to better accommodate generative artificial intelligence (gen-AI or AI), with a specific focus on protecting the rights of owners and authors of copyrighted works while still encouraging computational innovation.<sup>1</sup>

The consultation document specifically enquires about the use of copyright-protected works in text and data mining activities (TDM), authorship and ownership of works created by gen-AI, and how liability should be apportioned where works created by gen-AI result in copyright infringement. This report provides the context in which each of these issues arise under the current Canadian copyright regime and summarizes CIPPIC's stance on the inquiries posed by the consultation.

## 1.1. Summary of CIPPIC's Position

CIPPIC is a public interest clinic that specializes in technology law. Our goal is to advocate in the public interest for policy that promotes innovation, encourages respect for human rights, and responds to the needs of the wider public. These principles underlie the following recommendations that we offer in this submission:

- Gen-AI development remains a human-dependent process. Self-programming AI has not advanced to the point where models can generate desired outputs without the need for human input and specification. As a result of these computational constraints, humans should generally remain as the rights-holders for any produced works and liable for any infringement caused by Gen-AI outputs (Section 2).
- Where TDM is applied to copyrighted material for the purpose of building a training dataset for Gen-AI, there should be no claim for copyright infringement as long as the training data is not reproduced in any resulting generative output (Section 3).
- Authorship rights should remain with the human user responsible for prompting a Gen-AI output that falls within the scope of copyright protection. This approach allows copyright to extend to AI-generated works by focusing on the creativity and originality of the input, rather than judging the amount of creativity and originality in the creation of the algorithmic output (Section 4).
- Further clarity on how TDM and Gen-AI may result in copyright infringement is needed prior to the enactment of any legislation placing liability on a specific party for the infringing act. Namely, it must be determined whether the fault is that of the developers behind the Gen-AI for building and disseminating a model that permits substantial reproduction, or alternatively, the individual prompter for requesting an infringing output (Section 5).

---

<sup>1</sup> *Copyright Act*, RSC 1985, c C-42.

## 2. Technical Aspects of AI Development & Human Involvement

### 2.1. Context

The Canadian Government is seeking clarification on the general understanding of human involvement in the development of AI systems. The purpose on this inquiry is to determine how gen-AI models themselves may be seen as a separate legal entity under Canadian law, independent from the human counterparts responsible for developing the model. Such a perspective would permit for copyrights to vest in AI models, with the model itself being seen as the author or creator of a generated, copyrightable work.

While gen-AI has opened the door to AI involvement in the production of works, CIPPIC observes that human involvement remains. Authorship must vest in a human, if at all.

### 2.2. CIPPIC's Submissions

#### 2.2.1. *Human Involvement in AI Development*

The development of AI is human reliant. While self-programming AI systems have been theoretically formulated by many scholars, to date there has been no successful system of this kind due to current computational constraints. There has been success developing self-modifying AI systems using code-generating language models, meaning that the system is able to manipulate its own hyperparameters to improve its operation. Still, such a model does not involve actual AI self-programming but requires a human to develop and implement the model itself.<sup>2</sup>

Similarly, programs such as Codex (Open AI) are able to generate code from natural language inputs.<sup>3</sup> Like all current, publicly available generative AI systems, however, this system was built and made available by a human developer, and further requires prompting by a human user to produce the desired output. As a result, humans are still integral to the development of current AI systems.

#### 2.2.2. *The use of Gen-AI in the Legal Field*

CIPPIC is a public interest technology law clinic, meaning that our area of work primarily involves monitoring and intervening in policy issues and discussions arising at the intersection of law and advancing technologies. In the Canadian legal landscape, AI-assisted and AI-generated content can assist lawyers in reviewing contract formalities, generating memos and factums, as well as with research on specific legal topics as prompted by the user. Legal clients may also use generative AI-systems to ask for suggestions on how to approach a legal issue they are facing, though this should not be considered legal advice.

---

<sup>2</sup> Alex Sheng & Shankar Padmanabhan, "Self-Programming Artificial Intelligence Using Code-Generating Language Models" (2022) [unpublished, withdrawn submission archived at ICLR 2023 Conference], online: <<https://openreview.net/forum?id=SKat5ZX5RET>>.

<sup>3</sup> Wojciech Zaremba, Greg Brockman & OpenAI, "OpenAI Codex" (10 August 2021), online (blog): <<https://openai.com/blog/openai-codex>>.

## 3. The Impact of Text & Data Mining (TDM) on Copyright

### 3.1. Context

In the wake of widespread use of gen-AI platforms such as ChatGPT, fears have arisen regarding Canada's approach to copyright and the potential for infringement through text and data mining (TDM) activities.

In the gen-AI realm, TDM involves the process of scraping data from digital sources to create training datasets for the AI model to learn from. Authors have expressed concern over the ability for their works to be scraped and used in training datasets without authorization, specifically where the gen-AI may memorize and reproduce their works in a substantially similar manner in its generative output. In this sense, there are two primary ways that infringement can allegedly occur: first, through the act of completing TDM activities on copyrighted works to create a training dataset (an issue of input), and secondly, through developing a gen-AI model that has the capability to reproduce works used for training purposes (an issue of output).

The current copyright regime fails to address whether infringement arises in either of these scenarios. Focusing specifically on the first potential ground, however, CIPPIC is of the position that TDM activities alone should not be grounds for infringement, as it would result in undue limitations on innovation in gen-AI development. Furthermore, the government should be mindful of how imposing blanket liability for copyright infringement via TDM would affect both the vast alternative uses of TDM outside of the gen-AI field, as well as the balancing purpose of copyright. While there will be situations where TDM does indeed result in infringement (i.e., where the gen-AI is built for the purpose of reproduction), there will also be many scenarios where the TDM activity falls within the scope of fair dealing.

For these reasons, we call on the Canadian government to clarify the scope of liability for copyright infringement via TDM such that innovators and authors alike can better understand their rights and any limitations on such rights when copyright-protected work is used in AI development. This includes a discussion of how the current licensing regime may apply to the use of copyrighted works in training datasets, as well as the challenges that exist in implementing such an approach. As outlined below, our submission highlights the need for the government to regulate TDM with a holistic and cohesive view of its positive impacts across several sectors, such that any proposed solution strikes the necessary balance between author and creator rights and the public interest in innovation.

### 3.2. CIPPIC's Submissions

#### 3.2.1. *Necessary Clarification for TDM and its Impacts on Copyright in Canada*

Further clarity around copyright and TDM could shed light on the following issues:

- The nature of the copyrighted content being scraped for TDM purposes, and whether the *type* of content has implications for copyright infringement that may occur through TDM;
- How broadening authors' rights to capture TDM may shift copyright's balance and in so doing create risk and uncertainty for innovative activity (i.e., shifting to protection of ideas rather than expressions; violations of technological neutrality; unforeseen effects on non-technological applications of learning techniques); and
- How authors who would like to prevent their copyrightable digital works from being scraped for TDM purposes may do so effectively.

Such clarity would ensure innovation industries are aware of any copyright-imposed limitations on TDM activities, including whether TDM itself may implicate copyright infringement. The creative industry will benefit from gaining clarity on author rights and limitations on those rights when it comes to preventing their work from being used by TDM systems for use in AI-training datasets.

### **3.2.2. The Prevalence of TDM in Canadian Industries**

TDM activities are conducted in Canada and have been conducted since prior to the advent of modern generative AI programs. With origins in the late 1980s to early 1990s, TDM works to analyze large datasets and seek out patterns, trends, and insights. Importantly, TDM techniques are just one of many that can be employed by an AI system, meaning that TDM activities are a specific application and/or component of a support system for AI, rather than being equivalent to a generative AI program itself.

In Canada, TDM activities are currently observed across a wide variety of sectors, including academic research and literature analysis, business intelligence and marketing, as well as generative AI training and development through natural language processing. TDM is also used by the Canadian healthcare system for the analysis of electronic health records and genomic data, such that health professionals can discover patterns and potentially improve patient outcomes. Generally, TDM activities play a significant role in assessing the Canadian population and consumer patterns, making it a key component to informed decision-making, even when separated out of a generative or predictive AI model.

It is important to recognize that TDM describes a practice, not a technology. TDM can occur in analog form. Ultimately, it is a human practice. It does not necessarily implicate works. Indeed, TDM in the Generative AI context is just one application of a wider practice that has been a staple of innovative research and applications for some time now. Accordingly, any policy proposal to address this practise must consider the risks and implications of the change outside of the AI industry, and for practices involving or similar to TDM such as structuring and indexing information and innovating with technological systems that process information such as search engines, plagiarism detection, digitization initiatives and big data applications.

It is worth taking a moment to consider whether we are in the midst of a panic with the emergence of a powerful new technology. Any move to subject the development of AI to the controls and risks inherent to the copyright regime – and to its property-based system of exclusive rights, tariff structures, and robust remedy scheme – must proceed on the basis of an accurate understanding of how and when AI models interact with works. Equally, addressing the challenges of AI will involve an appreciation of the purposes of copyright law and the nature of author’s rights. As Justice Binnie famously observed in *Théberge v. Galerie d’Art du Petit Champlain Inc.*:

The *Copyright Act* is usually presented as a balance between promoting the public interest in the encouragement and dissemination of works of the arts and intellect and obtaining a just reward for the creator (or, more accurately, to prevent someone other than the creator from appropriating whatever benefits may be generated). [...]

The proper balance among these and other public policy objectives lies not only in recognizing the creator’s rights but in giving due weight to their limited nature.<sup>4</sup>

Any legislative amendment proposal should proceed only with a clear-eyed appreciation of its consequences for copyright’s virtuous balance.

---

<sup>4</sup> *Théberge v Galerie d’Art du Petit Champlain Inc*, 2002 SCC 34 at paras 30-31.



### **3.2.3. Challenges for Copyright Holders in Licensing Use of Works for TDM Purposes**

Canadian copyright holders often face challenges in licensing their works for TDM activities, in addition to facing challenges in preventing their copyright protected work from being used for TDM. The lack of clarity in Canadian law regarding how TDM may infringe copyright through unauthorized reproduction of works leaves rights holders with uncertainty as to their entitlement to licenses, the specific nature of their licensing rights, as well as the content and scope of potential licensing agreements. The further lack of direction on the applicability of fair dealing to TDM activities also introduces uncertainty as to when infringement occurs. More so, the diverse nature of TDM activities and works sought for TDM has led to a lack of standardization in licensing agreements, as well as undue complexity in determining licensing terms. This poses challenges for rights holders assessing fair compensation and defining the scope of licenses to use their works.

### **3.2.4. Current TDM Licensing Practices and its Imposed Challenges**

In Canada, there are various approaches to licenses for TDM activities. The most common form for publicly accessible data is terms of use agreements, which lay out the scope of permissible data use. Within these agreements, there may be specific permissions for text and data mining activities as well as conditions regarding how the results of the TDM may be used. For example, the terms of use may indicate that TDM activities are only permissible when completed for non-commercial purposes. Permissible use of the data may be dependent on the payment of a subscription or one-time fee to access the copyrighted material. For many of these sources, the individual seeking to conduct TDM agrees to the terms of use by performing the TDM activity itself, rather than agreeing to a negotiated license with the copyright holder.

Since Canadian copyright law fails to address text and data mining generally, those seeking to perform TDM activities must investigate the terms of use for each specific data source to ensure they have the necessary permissions to complete TDM. Similarly, those seeking to conduct TDM activities must ensure that their proposed use of the data aligns with what is actually and practically allowed by the agreement's terms. In the absence of a TDM provision from a terms of use agreement, express permission should be acquired from the copyright holder.

The current licensing scheme poses several challenges to those seeking to conduct TDM activities and copyright holders alike, as there is no clear, consistent approach to licenses for use. Terms of use agreements for TDM are employed by many largescale publishing companies, however, such terms are absent for smaller creators and copyright holders, leaving their works exposed to potentially unauthorized TDM. Similarly, those conducting TDM on a smaller scale face a resource and knowledge gap when compared to “big data” corporations, leaving them at a greater risk of violating license terms or being excluded from permission due to the potentially high costs of licenses, essentially pricing smaller innovators out of TDM and AI development.

Additionally, those seeking to complete TDM activities often face unduly limitations on data access, which hinders research and innovation pertaining to AI development. For example, the copyright holders who establish the licensing terms may limit the results of the TDM to specific word-limited extractions, rather than the whole of the results.<sup>5</sup> The inconsistencies among licensing schemes further limits accessibility, as the differences in terms of use may prevent comprehensive TDM-driven research and imposes unnecessary barriers in investigating and potentially paying for several different licenses.

---

<sup>5</sup> Peter McCracken & Emma Raub, “Licensing Challenges Associated with Text and Data Mining: How Do We Get Our Patrons What They Need?” (2023) 11:1 J Librarianship and Scholarly Communication eP15530 1 at 10–11.



Overall, there is uncertainty of entitlement to licenses under Canadian copyright law. Without clear guidelines on licensing, those seeking to complete TDM activities remain unaware of potential risks regarding copyright infringement, as well as their legal obligations and rights under licenses.

### **3.2.5. How the Copyright Act should view Text and Data Mining**

TDM activities should be permissible and not cause infringement as long as the training data is not reproduced in any resulting generative output. This means that the technology sector should be able to use copyrightable works in training datasets for AI in a manner that avoids infringement by reproduction.

CIPPIC's position is grounded in the balancing purpose of copyright protection, which lends itself to original expressions of an idea through the exercise of an author's skill and judgement, and not the idea itself.<sup>6</sup> As the Supreme Court of Canada observed, infringement can occur where a substantial portion of the work is reproduced, with the assessment being whether a substantial portion of the original work, as expressed by the author's skill and judgement, was copied.<sup>7</sup> Thus, whether a substantial portion of the work has been infringed is a question of quality rather than quantity and requires a holistic comparison of the works as a whole.<sup>8</sup>

Applying this to TDM activities, works subject to copyright protection would likely not be infringed through use in training data. TDM-based AI systems do not reproduce durable copies of these works, whether text or image-based. Rather, the AI is extracting information from unstructured (text mining) or structured data (data mining). It is through extracting this knowledge – essentially, deriving meta-data – that the AI system advances its capability to mimic human intelligence. Any technical and temporary reproductions arising from the analysis of training data for the purposes of constructing meta-data benefits from a number of user rights designed to facilitate innovation and its ensuing scientific, economic, creative, and artistic benefits. Fair dealing and the exception for temporary reproductions for technological processes both address these benefits. In other words, TDM activities alone do not give rise to an infringing reproduction, publication, or performance of the work.<sup>9</sup> Extending owner's rights to TDM would unduly extend protection to ideas, information, and data rather than the expression of the work.

### **3.2.6. Disclosure Obligations for use of Copyrighted Works in TDM and Gen-AI Training**

Considering the normative purpose of the *Copyright Act*, maintaining a balance between promoting the public interest in the encouragement and dissemination of works of the arts and intellect and obtaining a just reward for the creator, AI developers should disclose the use of copyright-protected content when used in the training of an AI system. Disclosure of the use of copyrighted content provides due acknowledgement to the original creator of the work even when the content itself is not being reproduced in any way, thereby upholding their moral rights to be associated with the work. At the same time, the public receives the benefit of AI systems with stronger and more accurate technological capacity, and AI developers are able to advance the technology through TDM activities without fear of unmerited legal claims. A way to address this normative position would be to specifically include TDM among those qualifying purposes of fair dealing that oblige the user to mention the sources and, if given, the author.

---

<sup>6</sup> *CINAR v Robinson*, 2013 SCC 73 at para 24 [*CINAR*]; *Copyright Act*, RSC 1985, c C-42 at s 5.

<sup>7</sup> *CINAR*, *ibid* at para 26.

<sup>8</sup> *CINAR*, *ibid* at para 26.

<sup>9</sup> *Copyright Act*, *supra* note 1 at s 3 & s 27(1).

### 3.2.7. *Remuneration for use of Copyrighted Works in TDM and Gen-AI Training*

Considering CIPPIC's position that there is no infringement caused by TDM activities alone, no remuneration would be required. However, if TDM activities are considered copyright infringement, a royalty akin to a typical licensing fee payable to the original author would align with technological neutrality principles, which require that the *Copyright Act* apply equally to different forms of media.<sup>10</sup> Such an approach recognizes that the use of copyrightable content in training data is equivalent to a license for using and reproducing the work in any other form.

Firms within the AI industry are entering into licenses with owners of copyrighted works to access those works for their use in TDM activities. These licenses reflect the value of the content towards enriching training data sets rather than recognizing existing obligations under copyright law. Structured data, for example, may have greater value than unstructured data for some purposes; similarly, well-edited data may have greater value than unedited social media data for some purposes. Licensing also reflects the ordinary reality of the value of avoiding expensive litigation.

CIPPIC supports the traditional tariff-setting practice of approximating tariff values to market conditions. The confidential nature of existing licensing practices complicates this approach.

If Canada were to adopt this approach to the application of TDM to copyright protected works, to minimize the harm to innovation and to ensure the AI market remains competitive and open to new entrants, Canada should ensure a remuneration model, not an exclusive rights model, for compensating authors. This system could also address authorship entitlement issues. It is difficult to associate TDM in copyright protected applications with value for individual works or authors except in the case of specific, author-centric libraries. In the vast majority of cases, publishers could claim the entirety of the value with no obligation to ensure Canadian authors obtain discrete benefits from the scheme.

### 3.2.8. *International TDM Approaches*

TDM is essential for AI development, and accordingly, Canadian copyright law ought to promote the activity. Other jurisdictions have adopted a variety of approaches to TDM, all with the common objective of accommodating TDM activities within existing copyright frameworks.

The US has yet to adopt any legislative or regulatory frameworks specific to AI and TDM activities. Rather, TDM is being decided in the courts under the fair use doctrine of existing copyright law. Section 107 authorizes the fair use of a copyrighted work, "including by reproduction ... for purposes such as criticism, comment, news reporting, teaching [...], scholarship, or research."<sup>11</sup> As applied to TDM, copying copyright protected works for the sole purpose of text and data mining has traditionally been considered a case of fair use by the technology sector in the US, usually justified by the transformative use factor of fair use under US copyright law.<sup>12</sup>

Japan updated its *Copyright Act* in 2018 to clarify that all users may analyze and understand copyrighted works for machine learning, create incidental copies of works, and use copyrighted works for data

---

<sup>10</sup> *Entertainment Software Association v. Society of Composers, Authors & Music Publishers of Canada*, 2012 SCC 34 at para 5; *Copyright Act*, *supra* note 1 at s 13.

<sup>11</sup> "U.S.C. Title 17 - COPYRIGHTS", online: <govinfo.gov/content/pkg/USCODE-2010-title17/html/USCODE-2010-title17-chap1-sec107.htm> at 17.

<sup>12</sup> "Text and data mining around the globe | Entertainment and Media Guide to AI | Perspectives | Reed Smith LLP", online: <reedsmith.com/en/perspectives/ai-in-entertainment-and-media/2023/06/text-and-data-mining-around-the-globe>.

verification. Thus, Japan allows researchers and private companies to carry out machine learning activities, including text and data mining.<sup>13</sup> Importantly, Japan differentiates between enjoyment and non-enjoyment of a copyrighted work. The purpose of Japan’s copyright law is to ensure compensation from those who want to enjoy the work. In the case of TDM, no one is enjoying the work; therefore, there is no copyright infringement. (Article 30-4, *Copyright Act* (Japan)).<sup>14</sup>

Europe implemented Articles 3 and 4 to Directive 2019/790 to allow TDM, albeit with restrictions.<sup>15</sup> However, it allows copyright holders to opt out of having their works mined. In 2014, the UK provided a copyright exception for TDM under s29A of the CDPA (1988).<sup>16</sup> The government had proposed to allow text and data mining “for any purpose” but withdrew the proposal in February 2023 after being challenged by the UK creative sector.<sup>17</sup>

## 4. Authorship and Ownership of AI Generated Works

### 4.1. Context

Gen-AI has the capability to produce works based on human inputs or “prompts.” As a result of this newfound proficiency, issues of authorship and ownership for works generated entirely or partially by AI have arisen, with the Government considering amending the *Copyright Act* to account for gen-AI’s potential role in authorship.

The primary legal issue is whether copyright can vest in gen-AI systems that can be prompted to create works. CIPPIC takes the position that as of the current state of the technology, copyright cannot and should not vest in the gen-AI but instead remain with the human responsible for prompting the creative output. While Gen-AI does have the capability to generate copyrightable works, its capability to do so is limited by human input. Many humans remain involved in the development and training of AI models and are further necessary to creating these works in their role as a prompter.

For the reasons below, any changes to the *Copyright Act* regarding the determination of first authorship in respect of AI-assisted or AI-generated works must align with the notion of AI’s human-dependency.

---

<sup>13</sup> “Japan amends its copyright legislation to meet future demands in AI”, (3 September 2018), online: *European Alliance for Research Excellence* <eare.eu/japan-amends-tdm-exception-copyright/>.

<sup>14</sup> Artha Dermawan, “Text and data mining exceptions in the development of generative AI models: What the EU member states could learn from the Japanese ‘nonenjoyment’ purposes?” n/a:n/a *The Journal of World Intellectual Property*, online: <onlinelibrary.wiley.com/doi/abs/10.1111/jwip.12285>, s 4.2-4.3.

<sup>15</sup> “EUR-Lex - 32019L0790 - EN - EUR-Lex”, online: <eur-lex.europa.eu/eli/dir/2019/790/oj> arts 3, 4 Doc ID: 32019L0790Doc Sector: 3Doc Title: Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (Text with EEA relevance.)Doc Type: LUsr\_lan: en.

<sup>16</sup> Expert Participation, “Copyright, Designs and Patents Act 1988”, online: <legislation.gov.uk/ukpga/1988/48/section/9>, s 29A.

<sup>17</sup> “UK withdraws plans for broader Text and Data Mining (TDM) copyright and database right exception”, (1 March 2023), online: *Intellectual property notes* <hsfnotes.com/ip/2023/03/01/uk-withdraws-plans-for-broader-text-and-data-mining-tdm-copyright-and-database-right-exception/>.

## 4.2. CIPPIC's Submissions

### 4.2.1. *Uncertainties surrounding Authorship and Ownership for Gen-AI Works*

The current copyright framework does not explicitly address AI-generated works, leading to ambiguities in cases where works are created with significant mixed human and AI involvement. The *Act* should be updated to explicitly state that authorship is the exclusive domain of humans. The primary benefit of such an amendment is to head off needless litigation and administrative burdens imposed by actors attempting to assert copyright authorship for algorithms, which, among other things, lack legal personhood to hold such rights.

Copyright theory, legal doctrines, and its underlying rationales, require a human author. The *Copyright Act* already assumes authors are human; otherwise, s. 6 of the *Act*, tying the term of copyright to the lifespan of the author, would be meaningless when considering AI.<sup>18</sup> For copyright to vest, a work requires the exercise of skill and judgment, and the work must not be so trivial that it could be characterized a purely mechanical exercise.<sup>19</sup>

As long as human input is required, AI should be viewed as a tool instead of an author. The originality and creativity elements will need to be judged on a case-by-case basis by examining the input (i.e., prompt) as well as the work itself. A generic input like “a picture of a cat” arguably lacks the originality needed for copyright to arise. The more specific and creative the human input, the stronger the case for copyright protection of the resulting work.

Outputs that lack original human input are unauthored and fall into the public domain.

### 4.2.2. *Amending Copyright Authorship and Ownership in light of Gen-AI*

Our position is that the human providing specific inputs to an AI should hold the copyright, especially when inputs are not generic, which aligns with the principles of existing copyright law. However, it would be beneficial for the Canadian Government to clarify these aspects to address the evolving landscape of AI-generated works. This approach maintains technology neutrality while ensuring that copyright continues to protect human creativity and expression.

### 4.2.3. *International Authorship and Ownership Approaches*

There are 3 broad national approaches addressing authorship of AI-generated works in copyright law.<sup>20</sup> First, the United States, Australia, and most continental European countries require human creativity in copyright law and does not extend copyright protection to AI-generated works.<sup>21</sup>

Secondly, the United Kingdom, New Zealand, South Africa, and India award authorship through legislation to the human that arranged the work, and broadly permits fully autonomous or sentient AI to author

---

<sup>18</sup> Dale Smith, “Creativity in the Age of AI” (6 Aug 2021), online: <nationalmagazine.ca/en-ca/articles/law/in-depth/2021/creativity-in-the-age-of-ai>.

<sup>19</sup> *CCH v Law Society of Upper Canada*, 2004 SCC 13 at para 16 [CCH].

<sup>20</sup> “WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): Third Session”, online: <wipo.int/meetings/en/details.jsp?meeting\_id=59168>, see PDF Summary: <wipo.int/edocs/mdocs/mdocs/en/wipo\_ip\_ai\_3\_ge\_20/wipo\_ip\_ai\_3\_ge\_20\_inf\_5.pdf>.

<sup>21</sup> *Thaler v Perlmutter*, case No. 1:22-cv-01564, (D.D.C. 8/18/23) at 2; “US Court Decides There is No Copyright in AI-Generated Works - What About Canada? | Cassels.com”, online: *Cassels* <cassels.com/insights/us-court-decides-there-is-no-copyright-in-ai-generated-works-what-about-canada/>.

works (see s. 9(3), *Copyright, Designs, and Patents Act of 1988* (CDPA) – UK). The United Kingdom allows a copyright to subsist in AI-generated works by attributing authorship of the works to the human, corporate, or AI machine author that simply arranged the final copyrighted work. Thus, the United Kingdom relies on “skill and labour” or sweat of brow to determine who arranged the work (CPDA s. 9(3)).<sup>22</sup>

Finally, China and Japan use the judicial system to incrementally expand upon existing legislation by attributing copyright authorship to human programmers and companies that create code dictating AI’s creative decisions and by declining to extend authorship to AI.<sup>23</sup> For example, in *Shenzhen Tencent v. Shanghai Yingxun* (2019), the Chinese judiciary extended copyright protection to AI-generated works and attributed authorship in the final work to the human author or organization that created the AI. In *Gao Yang et al. v. Golden Vision* (2020), high-altitude photographs taken automatically by AI merited copyright protection because although humans did not click the shutter-release button to take the photograph (the AI made this decision), humans were solely responsible for making creative decisions that influenced the high-altitude photographs, such as the shooting angle, video recording mode, and video display format.<sup>24</sup>

Our suggestion is that Canada follows the Chinese approach to authorship by attributing copyright authorship to the humans or corporations that create the code or prompt dictating the AI’s output. This approach allows copyright to extend to AI-generated works by focusing on the creativity and originality of the input, rather than judging the amount of creativity and originality in the creation of the output. It also avoids creating sui generis rights or legal fictions to accommodate AI-generated creations. This approach is more open to considering AI’s role as a tool in the creative process.

## 5. Liability for Infringement by Gen-AI

### 5.1. Context

As discussed above, there is vast concern regarding Canada’s application of the copyright regime to gen-AI, specifically in terms of liability for copyright infringement. There are two primary ways that infringement can allegedly occur: first, through the act of completing TDM activities on copyrighted works to create a training dataset (an issue of input), and secondly, through developing a gen-AI model that has the capability to reproduce works used for training purposes (an issue of output).

Here, we focus on *who* would be liable for such infringement and *when* liability should arise, with CIPPIC taking the stance that copyright law should remain human-centric in its determination of liability until a more cohesive framework is developed. Such a framework is necessitated by the barriers posed by gen-AI regarding how models actually use copyright-protected training data to generate works, as well as the responsibility of prompters in requesting infringing outputs. As detailed below, the key issue that must be answered is *where* the infringement arises – is it the fault of the AI-developers for building a model that has the capability to infringe, or the fault of individual users for requesting an infringing output?

---

<sup>22</sup> Although note that recently Lord Justice Arnold in *THJ Systems Limited & Anor v Daniel Sheridan & Anor* [2023] *EWCA Civ 1354* at para 23 identified the correct test for originality under the statute was whether the work was the “author’s own intellectual creation”, and no longer “skill and labour”.

<sup>23</sup> ZHOU Bo, “Artificial Intelligence and Copyright Protection --Judicial Practice in Chinese Courts”.

<sup>24</sup> “Does China Back Copyrights for Automatic Photos from a Hot-Air Balloon? - China Justice Observer”, online: <[chinajusticeobserver.com/a/does-china-back-copyrights-for-automatic-photos-from-a-hot-air-balloon](http://chinajusticeobserver.com/a/does-china-back-copyrights-for-automatic-photos-from-a-hot-air-balloon)>.

Without legislative or judicial directive on such issues, AI-developers and creators alike will remain uncertain of their legal rights, limitations, and obligations surrounding the use of copyright-protected works both in gen-AI training and any resulting outputs. Considering the exponential growth in use of gen-AI by the Canadian public, such clarity is needed as soon as possible.

## **5.2. CIPPIC's Submissions**

### **5.2.1. Applying the Current Infringement Framework to Infringing, AI-Generated Works**

As described above, whether a substantial portion of the work has been infringed is a question of quality rather than quantity and requires a holistic comparison of the works as a whole.<sup>25</sup> In other words, infringement can be found for both literal and non-literal copying where the substantial quality of the work is reproduced.<sup>26</sup> As the Supreme Court has ruled, an assessment of substantial copying focuses on whether copied features constitute a substantial part of the original work of the author; thus, the alteration of copied features or its integration into a notably different work may not preclude an infringement claim if a substantial quality of the work has been copied.<sup>27</sup>

As a result, generative AI systems may face infringement claims if their outputted works copy a substantial portion of the quality of the original work that the system was trained on (i.e., substantial reproduction of the copyright protected training data). This will be extremely difficult to monitor considering the expansiveness of most AI training data sets, as elements from a multitude of different works may be combined to produce a generative output. Infringement may be clearer when the prompter asks the AI system to generate an output addressing the specific expression of an author, artist, musician, or other copyright-protected creator. The law currently accommodates this issue by holistically assessing each alleged infringement on a case-by-case basis; copyright infringement is a matter of degree, nuance, and context, and whether a substantial part of a work has been copied is a flexible and fact-specific notion.<sup>28</sup>

Due to the nuanced nature of this test, concerns are likely to focus on where an AI-generated work, trained by TDM activities, including copyright-protected expression, is being commercialized. In this situation, creators may raise issues related both to their moral and economic rights. For moral rights, authors may raise concerns related to how the integrity of their work has been manipulated by the AI system, as well as the loss of association with a substantially copied derivative of their original work. Considering that economic rights include the right to authorize reproductions, copyright owners may also raise concerns when another party exercises any of the exclusive rights associated with their work without consent.<sup>29</sup>

### **5.2.2. Barriers to Establishing Infringement by Gen-AI: Access to the Original Work**

The vast majority of modern-day AI uses deep learning. Deep learning is a subset of AI and uses artificial neural networks to mimic the human brain's learning process. The one pitfall of deep learning is that generated outputs are developed using black-box algorithms, meaning that users are unable to see how the deep learning system actually makes its decisions. While we understand the training data inputted to the AI model and can observe the outputs created, we have no way of knowing how the AI actually came to that output. The only indicator we likely would have is the user-generated code used as the original scaffolding for the AI's learning patterns. In other words, considering both the black-box algorithm issue

---

<sup>25</sup> *CINAR*, *supra* note 6 at para 26.

<sup>26</sup> *CINAR*, *supra* note 6 at para 27.

<sup>27</sup> *CINAR*, *supra* note 6 at para 39.

<sup>28</sup> *CINAR*, *supra* note 6 at paras 26 & 40.

<sup>29</sup> *Copyright Act*, *supra* note 1 at s 3; *Théberge v Galerie d'Art du Petit Champlain Inc*, 2002 SCC 34 at para 12.



and the expansive breadth of datasets, there is no way for us to assess whether or not the AI accessed a specific copyright-protected work. Our knowledge is limited to the inclusion or exclusion of the copyright-protected work in the original dataset.

However, the core principle of how generative AI models function does not implicate copyright infringement: AI researchers do not design AI systems to reproduce training data; they design them to abstractly “learn” from training data. Training data influences the algorithm, but the algorithm does not reproduce the training data.

In rare cases, generative AI models reproduce copyright-protected expression. A precise understanding of how, why, and when this occurs should predicate any policy conclusions the Canadian government draws from this phenomenon. CIPPIC’s understanding of this phenomenon suggests that it arises from specific technological phenomena combined with user-specific prompts. AI researchers are better able to describe the scope and limits of this phenomenon.

Larger concerns arise from copyright protection for the reproduction of abstract concepts of expression, such as fictional characters.<sup>30</sup> Caselaw will, over time, provide greater clarity over the limits of expression of abstract ideas that reproduce expression subject to copyright protection. Developers of AI systems will need to consider mechanisms for identifying where such occurrences are likely to emerge.

### **5.2.3. Commercial Use & Risk Mitigation for Infringing AI-Generated Works**

How the Canadian *Copyright Act* applies to AI-generated outputs is unclear, leaving businesses and other organizations unsure of liability for copyright infringement through AI-applications. As a result, Canadian enterprises are adopting a patchwork of risk-mitigation strategies. Examples include restricting training data for AI software such that all data used is either licensed or public-domain; corporate indemnification clauses to protect end users from infringement claims where their use of AI was within the scope of the software’s terms and conditions; and author-applied tags to label their works as being non-TDM friendly.

Some enterprises have combined these approaches to create a robust framework protecting AI application users from infringement liability. For example, Adobe Firefly, which uses generative AI to alter an image based on user-inputted text prompts, has co-founded and implemented the Coalition for Content Provenance and Authenticity (C2PA) and the Content Authenticity Initiative (CAI).<sup>31</sup> The C2PA and CAI work together to permit publishers, creators, and consumers to trace the origin of different types of media using an open technical standard. These tools allow users to add Content Credentials that allow the creator to indicate that generative AI was used in the production of the work.<sup>32</sup> Information about the specific AI models used can be traced by the user, thereby helping to increase transparency and prevent the spread of misinformation regarding the use of generative AI in creative works.<sup>33</sup>

Adobe has also made efforts to ensure that Firefly’s commercial character does not lead to infringement claims by training the AI model on licensed (Adobe Stock) and public domain content, while not training the AI using any subscribers’ personal content.<sup>34</sup> Adobe Stock contributors whose works have been used to train Adobe Firefly are eligible for a “Firefly bonus compensation plan.” This plan represents a license

---

<sup>30</sup> See *CINAR*, *supra* note 6, for example.

<sup>31</sup> “Overview - C2PA”, online: <c2pa.org/>; “Content Authenticity Initiative”, online: *Content Authenticity Initiative* <contentauthenticity.org>.

<sup>32</sup> *Ibid.*

<sup>33</sup> *Ibid.*

<sup>34</sup> *Ibid.*



agreement whereby contributors are paid for the use of their work as training data in the Firefly AI software.<sup>35</sup> The purpose of Adobe’s approach was to eliminate the potential for copyright infringement claims, with Adobe going so far as to offer intellectual property indemnification for any legal issues arising from its use.<sup>36</sup> This indemnification clause provides that as long as the user has used the Firefly product in accordance with the terms and conditions, Adobe will compensate the individual for any IP-related legal claims that may arise.<sup>37</sup> As a result, any generated work incorporating Firefly AI within the scope of its authorized terms will not be subject to personal copyright infringement liability.

#### **5.2.4. Uncertainties on Liability for Copyright Infringement by an AI-Generated Work**

Further clarity is required in several areas related to AI systems generally, but specifically for generative AI systems and the works they produce:

1. Greater clarity on what constitutes a “substantial part” of a work is necessary for determining the liability of AI-generated works for infringement. There must be further clarification on the protection of artistic styles and whether that is considered as an unprotected idea or a protected expression of skill and judgement.
2. Greater clarity is required regarding *who* would be liable for copyright infringement in these circumstances. Considering that copyright authorship requires an original expression through an exercise of skill and judgement, the legislature must provide clarity on whether or not an AI system can truly fulfil these criteria. At its core, AI is a mathematical and statistical computer science tool that analyzes training data for correlations and patterns, and then it uses those patterns to generate a predictive output. Essentially, the AI is regurgitating and recombining data to create a novel output. The issue of whether this constitutes an exercise of skill and judgement, such that the AI system is an author of an original, expressive work, must be resolved by the legislature. If it does, the AI itself would be liable, which would require a novel remedy of some kind. If it does not, is the human prompter who generated the specific output liable? Or does liability fall onto the corporation or individual who coded the AI system itself?

Such clarity need not originate with legislative initiatives. In any event, special legislative amendments that depart from the usual copyright rules of liability for specific industries or technologies would violate the law’s neutrality. Experience has shown that such solutions are short-lived as the pace of the market and innovation inevitably leaves them behind. Industry standards, consensus best practices documents, and litigation can all contribute to providing greater certainty around AI innovations. The government can play a role in facilitating multi-stakeholder initiatives that can lend themselves to this end.

International perspectives on AI authorship and liability, described above under section 4.1.3, could be helpful in establishing the nature of liability where Gen-AI results in copyright infringement.

---

<sup>35</sup> “Firefly FAQ for Adobe Stock Contributors”, online: <[helpx.adobe.com/ca/stock/contributor/help/firefly-faq-for-adobe-stock-contributors.html](https://helpx.adobe.com/ca/stock/contributor/help/firefly-faq-for-adobe-stock-contributors.html)>.

<sup>36</sup> “Adobe Firefly Product Description: Firefly IP Indemnification”, online: <[helpx.adobe.com/ca/legal/product-descriptions/adobe-firefly.html](https://helpx.adobe.com/ca/legal/product-descriptions/adobe-firefly.html)>.

<sup>37</sup> *Ibid.*

## 6. Additional Considerations

The following section outlines further concerns raised by CIPPIC regarding copyright and gen-AI in our submission to the Government’s consultation.

### 6.1. CIPPIC’s Submissions

#### 6.1.1. Vicarious Liability

As with any neutral technology that interacts with works, users can accidentally or deliberately produce AI outputs that infringe copyright. This raises the question of whether proprietors of AI systems used in this way may be vicariously liable for the infringements of its users. Vicarious liability for copyright infringement arises through (a) the authorization of an infringing act<sup>38</sup> or (b) the provision of a service primarily for the purpose of enabling acts of copyright infringement.<sup>39</sup>

Whether a party “authorized” infringement is a question of fact, looking to whether the alleged authorizer sanctioned, approved, or countenanced the infringement.<sup>40</sup> A court may infer authorization from the facts, meaning that both positive acts and sufficient indifference or passivity to known infringement can lend grounds for a secondary infringement claim.<sup>41</sup> Notably, liability for authorizing infringement does not occur where a person authorizes the mere use of technology that a primary infringer could use to infringe copyright.<sup>42</sup> Courts look to the knowledge the authorizer possessed of the infringing acts, and the degree of control the alleged authorizer exercised over the primary infringer.

Authorization liability opens the door to liability for programmers, providers, and users of AI systems who prompt the generation of an infringing work. Regardless of the identity of the author of the infringing AI-generated work, an expansive approach to authorization could extend liability throughout the chain of people responsible for building the AI, training it, and prompting the work’s creation.<sup>43</sup> If a developer assembles a training data library of copyright-protected material, the distribution of that training data to others could give rise to an authorization claim. The programmers who train generative AI on such data sets and build the AI system such that it can reproduce a substantial, infringing portion of copyright-protected work could theoretically also face liability (see section 5.2.1, above).

The notion that infringement is not authorized where only equipment that could result in infringement is provided further complicates vicarious liability, as this rule indicates that only prompters should be liable for requesting the generation of an infringing work by the model. It must be clarified whether the issue is A) the potential for the AI model to reproduce a substantially similar output to its copyright protected training data or B) the ability for users to prompt such an output.

The manner in which courts have construed the authorization right should prove satisfactory in addressing such risks. AI services do not exercise control over any primary infringer producing outputs that infringe copyright. AI is neutral technology, and only in exceptional circumstances would the ordinary use case implicate a copyright infringement. Similarly, authorization does not require intervention to stop infringements on the part of an otherwise neutral by-stander, including the purveyor of technology used

---

<sup>38</sup> *Copyright Act*, *supra* note 1 at s 3(1).

<sup>39</sup> *Copyright Act*, *supra* note 1 at s. 27(2.3).

<sup>40</sup> *CCH*, *supra* note 19 at para 38.

<sup>41</sup> *CCH*, *supra* note 19 at para 38.

<sup>42</sup> *CCH*, *supra* note 19 at para 38.

<sup>43</sup> Carys Craig, “AI and Copyright” in Florian Martin-Bariteau & Teresa Scassa, eds, *Artificial Intelligence and the Law in Canada* (Toronto: Lexis Nexis, 2021) ch 1 at 15.

by another to infringe copyright. Copyright does not violate the liberty principle that underlies much of Canadian law – the law does not require Canadians to police the actions of our neighbours.

Expansive interpretation of the authorization provisions of the *Act* could result in the potential for broad vicarious infringement claims against all parties involved in the development of an AI system. For example, in *Voltage Holdings, LLC v. Doe #1*, the plaintiff has sought liability against internet subscribers on the basis of deemed knowledge of infringing acts, alleged control over the point of internet access, and an alleged failure to stop the infringements.<sup>44</sup> Should courts abandon the traditional control test (that looks to the legal and personal relationship between allegedly infringing actors) in favour of assumed technological control, authorization liability could become a significant risk for AI actors. Similarly, if the courts were to overturn long-standing precedent and import into authorization a duty to police or intervene in copyright wrongs, this too would raise red flags. It is worth commenting that radical policy shifts in the scope and reach of authorization would affect far more than just purveyors of AI services and should provoke a legislative reaction. Furthermore, the *Act*'s prohibition on the provision of infringement enablement services should prove adequately focused on bad actors to avoid its inadvertent deployment against content-neutral, general application of AI services.

### 6.1.2. Rights Management Information

The *Copyright Act*'s prohibition on altering or removing electronic rights management information (“RMI”) associated with an electronic copy of a work ought not to prove a concern for AI entrepreneurs.<sup>45</sup> This is so for both TDM activities and with respect to outputs. RMI liability requires a claimant meet a “triple knowledge” criteria: to be liable, a defendant must (1) “knowingly remove or alter” RMI, (2) without the consent of the copyright owner, and (3) in circumstances where the defendant “knows or should have known” that the removal or alteration “will facilitate or conceal any infringement of the owner’s copyright.”<sup>46</sup> The particularity of these knowledge requirements, and their inapplicability to cases involved fair dealing or other exceptions to infringement – greatly limit the potential of RMI tampering claims to frustrate AI research and commercial applications.

## 7. Conclusion

While gen-AI is a newly emerging technology, it has quickly become a fundamental aspect of society and the Canadian marketplace. Considering its implications on copyrighted materials, reforming the Canadian copyright regime is a necessary step in resolving one aspect of the multitude of legal challenges raised by gen-AI. Ultimately, gen-AI will have substantial effects on every economic sector, meaning that a coherent approach across the Canadian legal landscape is necessary. Applied to copyright, this means that application of the *Copyright Act* to this novel technology must align with the purpose of the *Act* and assist in striking the balance between creator rights and the public interest in encouraging innovation. As illustrated, CIPPIC’s position attempts to bring the law closer to this purpose, while simultaneously recognizing the human-reliant nature of currently available gen-AI technologies.

---

<sup>44</sup> *Voltage Holdings, LLC v Doe*, 2023 FCA 194.

<sup>45</sup> *Copyright Act*, *supra* note 1 at s 42.22(1).

<sup>46</sup> *Ibid.*